

# CERTIFYING CONVERGENCE OF LASSERRE'S HIERARCHY VIA FLAT TRUNCATION

JIAWANG NIE

**ABSTRACT.** Consider the optimization problem of minimizing a polynomial function subject to polynomial constraints. A typical approach for solving it is applying Lasserre's hierarchy of semidefinite programming relaxations, based on either Putinar's or Schmüdgen's Positivstellensatz. A practical question in applications is: how to certify its convergence and get minimizers? In this paper, we propose *flat truncation* as a general certificate for this purpose. Assume the set of global minimizers is nonempty and finite. Our main results are: i) Putinar type Lasserre's hierarchy has finite convergence if and only if flat truncation holds, under some general assumptions, and this is also true for the Schmüdgen type one; ii) under the archimedean condition, flat truncation is asymptotically satisfied for Putinar type Lasserre's hierarchy, and similar is true for the Schmüdgen type one; iii) for the hierarchy of Jacobian SDP relaxations, flat truncation is always satisfied. The case of unconstrained polynomial optimization is also discussed.

## 1. INTRODUCTION

Given polynomials  $f, g_1, \dots, g_m$ , consider the optimization problem

$$(1.1) \quad \begin{cases} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & g_1(x) \geq 0, \dots, g_m(x) \geq 0. \end{cases}$$

Let  $K$  be its feasible set, and  $f_{\min}$  be its global minimum. The  $k$ -th Lasserre's relaxation [6] for solving (1.1) is ( $k$  is also called the order)

$$(1.2) \quad \max \quad \gamma \quad \text{s.t.} \quad f(x) - \gamma \in Q_k(g).$$

Here, the set  $Q_k(g)$  denotes the  $k$ -th truncated quadratic module generated by the tuple  $g := (g_1, \dots, g_m)$  (for convenience, denote  $g_0 \equiv 1$  and say a polynomial is SOS if it is a sum of squares of other polynomials):

$$Q_k(g) := \left\{ \sum_{i=0}^m g_i \sigma_i \mid \sigma_i \text{ is SOS, } \deg(g_i \sigma_i) \leq 2k \text{ for every } i \right\}.$$

The relaxation(1.2) is equivalent to a semidefinite programming (SDP) problem, and thus could be solved efficiently by numerical methods like interior point type algorithms. Let  $f_k$  be the optimal value of (1.2) for a given order  $k$ . Clearly, every  $f_k \leq f_{\min}$  and the sequence  $\{f_k\}$  is monotonically increasing. Under the archimedean condition (i.e., there exists  $\phi \in Q_\ell(g)$  for some  $\ell$  such that the inequality  $\phi(x) \geq 0$  defines a compact set in  $x$ ), Lasserre [6] proved a fundamental

---

1991 *Mathematics Subject Classification.* 65K05, 90C22.

*Key words and phrases.* flat truncation, Lasserre's relaxations, quadratic module, preordering, semidefinite programming.

result:  $f_k \rightarrow f_{\min}$  as  $k \rightarrow \infty$ . An estimation of its convergence rate is given in [12]. The sequence of (1.2) as  $k \rightarrow \infty$  is also called *Lasserre's hierarchy* in the literature. As demonstrated by extensive numerical experiments, it occurs quite a lot that  $f_k = f_{\min}$  for a finite order  $k$  in applications. If this happens, we say Lasserre's hierarchy has *finite convergence*. This raises a very practical question: since  $f_{\min}$  is typically unknown, how do we certify its finite convergence if it happens? If it is certified, how do we get minimizers? A frequently used sufficient condition for this purpose is *flat extension* introduced by Curto and Fialkow (cf. [3]), but it is a strong condition that might not be satisfied, i.e., it is not necessary. To the author's best knowledge, there is very few work on proving a general certificate for checking finite convergence of Lasserre's hierarchy, and the question is almost completely open. The motivation of this paper is to address this issue. Our main result is that a more general condition called *flat truncation* could typically serve as such a certificate.

**1.1. Background.** Typically, to extract a global minimizer, one needs to consider the dual optimization problem of (1.2) whose description uses localizing matrices. Define degree integers

$$(1.3) \quad d_i = \lceil \deg(g_i)/2 \rceil, \quad d_g = \max\{1, d_1, \dots, d_m\}, \quad d_f = \lceil \deg(f)/2 \rceil.$$

(Here  $\lceil a \rceil$  denotes the smallest integer that is greater than or equal to  $a$ .) Let  $y$  be a sequence indexed by  $\alpha := (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$  ( $\mathbb{N}$  is the set of nonnegative integers) with  $|\alpha| := \alpha_1 + \dots + \alpha_n \leq 2k$ , i.e.,  $y$  is a *truncated moment sequence (tms)* of degree  $2k$ . Denote by  $\mathcal{M}_{2k}$  the space of all tms whose degrees are  $2k$ . A tms  $y \in \mathcal{M}_{2k}$  defines a Riesz functional  $\mathcal{L}_y$  on  $\mathbb{R}[x]_{2k}$  (the space of real polynomials in  $x := (x_1, \dots, x_n)$  with degrees at most  $2k$ , and denote  $\mathbb{R}[x] = \sum_i \mathbb{R}[x]_i$ ) as

$$\mathcal{L}_y \left( \sum_{\alpha} p_{\alpha} x_1^{\alpha_1} \cdots x_n^{\alpha_n} \right) = \sum_{\alpha} p_{\alpha} y_{\alpha}.$$

For convenience, denote  $x^{\alpha} := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ ; if  $p(x) = \sum_{\alpha} p_{\alpha} x^{\alpha}$ , we also denote

$$\langle p, y \rangle = \mathcal{L}_y(p).$$

The  $k$ -th *localizing matrix*  $L_h^{(k)}(y)$  generated by a polynomial  $h$  and a tms  $y \in \mathcal{M}_{2k}$  is a symmetric matrix satisfying (denote  $d_h = \lceil \deg(h)/2 \rceil$ )

$$p^T L_h^{(k)}(y) p = \mathcal{L}_y(h(x)p(x)^2), \quad \forall p(x) \in \mathbb{R}[x]_{k-d_h}.$$

On the left hand of the above,  $p$  denotes the coefficient vector of the polynomial  $p(x)$ . When  $h \equiv 1$ ,  $L_h^{(k)}(y)$  is called a *moment matrix* and is denoted as

$$M_k(y) := L_1^{(k)}(y).$$

The columns and rows of  $L_h^{(k)}(y)$ , as well as  $M_k(y)$ , are indexed by integral vectors  $\alpha \in \mathbb{N}^n$  with  $|\alpha| \leq k - d_h$ .

The dual optimization problem of (1.2) is (cf. [6, 8])

$$(1.4) \quad \begin{cases} \min_{y \in \mathcal{M}_{2k}} & \langle f, y \rangle \\ \text{s.t.} & L_{g_i}^{(k)}(y) \succeq 0 \ (i = 0, 1, \dots, m), \ \langle 1, y \rangle = 1. \end{cases}$$

In the above,  $X \succeq 0$  means a matrix  $X$  is positive semidefinite. Let  $f_k^*$  be the optimal value of (1.4) for order  $k$ . By weak duality,  $f_k^* \leq f_k$  for every  $k$ . If  $K$  has nonempty interior, then (1.4) has an interior point, (1.2) achieves its optimal value

and  $f_k^* = f_k$ , i.e., there is no duality gap between (1.2) and (1.4) (cf. [6]). Clearly, every  $f_k^* \leq f_{\min}$ , and the sequence  $\{f_k^*\}$  is also monotonically increasing. We refer to Lasserre's book [8] and Laurent's survey [10] for related work in this area.

Suppose  $y^*$  is an optimizer of (1.4). If the *flat extension condition* (cf. [3])

$$(1.5) \quad \text{rank } M_{k-d_g}(y^*) = \text{rank } M_k(y^*)$$

holds ( $d_g$  is from (1.3)), then we can extract  $r = \text{rank } M_k(y^*)$  global optimizers for (1.1). By Theorem 1.1 of Curto and Fialkow [3], if  $y^*$  is feasible for (1.4) and (1.5) is satisfied, then  $y^*$  admits a unique  $r$ -atomic measure supported on  $K$ , i.e., there exist  $(\lambda_1, \dots, \lambda_r) > 0$  and  $r$  distinct points  $v_1, \dots, v_r \in K$  such that

$$(1.6) \quad y^* = \lambda_1[v_1]_{2k} + \dots + \lambda_r[v_r]_{2k}.$$

Here, for  $x \in \mathbb{R}^n$ ,  $[x]_{2k}$  denotes the vector defined as

$$[x]_{2k} := [1 \quad x_1 \quad \dots \quad x_n \quad x_1^2 \quad x_1x_2 \quad \dots \quad x_n^{2k}]^T.$$

In (1.6), the constraint  $\langle 1, y^* \rangle = 1$  implies  $\lambda_1 + \dots + \lambda_r = 1$ . If (1.5) holds, then  $f_k = f_{\min}$  (suppose there is no duality gap between (1.2) and (1.4)), all  $v_1, \dots, v_r$  are global minimizers of (1.1), and they could typically be obtained by solving some SVD and eigenvalue problems, as shown by Henrion and Lasserre [5]. Generally, (1.5) is a sufficient but not necessary condition for checking finite convergence of Lasserre's hierarchy.

For a tms  $z \in \mathcal{M}_{2t}$ , we say  $z$  is *flat with respect to  $g$*  if  $z$  is feasible in (1.4) for  $k = t$  and satisfies  $\text{rank } M_{t-d_g}(z) = \text{rank } M_t(z)$  (cf. [3]). If the tuple  $g$  is clear in the context, we just simply say  $z$  is flat.

**1.2. Flat truncation.** To get a minimizer of (1.1) from an optimizer  $y^*$  of (1.4), the flat extension condition (1.5) would be weakened. For instance, if  $y^*$  has a *flat truncation*, i.e., there exists an integer  $t \in [\max\{d_f, d_g\}, k]$  such that

$$(1.7) \quad \text{rank } M_{t-d_g}(y^*) = \text{rank } M_t(y^*),$$

then we would still extract  $r := \text{rank } M_t(y^*)$  minimizers for (1.1). Let  $z = y^*|_{2t}$  (denote  $y^*|_{2t} := \{y_\alpha^* : |\alpha| \leq 2t\}$ ) be the truncation. Then, the tms  $z$  is flat, and there exist  $r$  distinct points  $u_1, \dots, u_r \in K$  and scalars  $\lambda_1, \dots, \lambda_r$  such that

$$z = \lambda_1[u_1]_{2t} + \dots + \lambda_r[u_r]_{2t},$$

$$\lambda_1 > 0, \dots, \lambda_r > 0, \quad \lambda_1 + \dots + \lambda_r = 1.$$

Clearly,  $\langle f, z \rangle = \langle f, y^* \rangle = f_k^* \leq f_{\min}$  (note  $2t \geq \deg(f)$ ) and

$$\lambda_1 f(u_1) + \dots + \lambda_r f(u_r) = \langle f, z \rangle = \langle f, y^* \rangle \leq f_{\min}.$$

Since every  $f(u_i) \geq f_{\min}$ , each  $u_i$  must be a global minimizer of (1.1). Hence, if an optimizer of (1.4) has a flat truncation, then  $f_k^* = f_{\min}$ , and  $f_k = f_{\min}$  if there is no duality gap between (1.2) and (1.4).

The rank condition (1.7) was used in [8, 10] as a sufficient condition to verify exactness of Lasserre's relaxations in polynomial optimization and generalized problems of moments. When the feasible set  $K$  is defined by polynomial equalities  $f_1(x) = \dots = f_s(x) = 0$  and possibly polynomial inequalities, if the equations have finitely many zeros, then Lasserre's hierarchy has finite convergence (cf. [9, §3] or [10, §6.5]), and for  $k$  big enough every tms  $y$  that is feasible for (1.4) has a flat truncation (cf. [7, Prop. 4.6] or [10, Theorem 6.20]). For general polynomial optimization, there are no such similar results in the existing literature.

As we have seen earlier, for an optimizer of (1.4) to have a flat truncation, a necessary condition is  $f_k = f_{\min}$  (suppose (1.2) and (1.4) has no duality gap). Thus, flat truncation is generally a sufficient condition for Lasserre's hierarchy to have finite convergence. Thus, one is wondering whether flat truncation is also necessary: if Lasserre's hierarchy converges in finitely many steps, does *every*<sup>1</sup> optimizer of (1.4) have a flat truncation? If so, flat truncation could be used as a certificate for checking its finite convergence and some minimizers of (1.1) could also be obtained. This issue was addressed very little in the field.

Another important issue in applications is about asymptotic convergence of Lasserre's hierarchy. Under the archimedean condition, we know  $f_k \rightarrow f_{\min}$  as  $k \rightarrow \infty$  (cf. [6]), but the convergence to the set of minimizers is not known very well. When (1.1) has a unique global minimizer  $u^*$ , Schweighofer [20] showed an important result: the subvector consisting of linear moments of a nearly optimizer of (1.4) converges to  $u^*$  as  $k \rightarrow \infty$ . However, for the general case that (1.1) has more than one global minimizer, there are no such similar results in the existing work, to the author's best knowledge. It is possible that  $f_k = f_k^* < f_{\min}$  for every  $k$ , i.e., Lasserre's hierarchy may have no finite convergence, as implied by Scheiderer's work [17]. In such cases, one should not expect any minimizer of (1.4) to have a flat truncation. However, how about the limit points of truncations of minimizers of (1.4) as  $k \rightarrow \infty$ ? Is every such limit point flat, or does it have a flat truncation? There is very few work on this issue.

**1.3. Contributions.** This paper focuses on proving a certificate for checking convergence of Lasserre's hierarchy. We assume (1.1) has a nonempty set  $S$  of global minimizers and its cardinality  $|S|$  is finite, which is generally true. Our main result is that flat truncation would generally serve as such a certificate.

First, we study how to certify finite convergence. For Lasserre's hierarchy of (1.2) (it is also called a Putinar type one), we show that: it has finite convergence if and only if every minimizer of (1.4) has a flat truncation when  $k$  is sufficiently large, under some general conditions (see Assumption 2.1). For Schmüdgen type Lasserre's hierarchy, which is a refined version of (1.2) by using cross products of  $g_j$  (see (2.6)), we also show that generally it has finite convergence if and only if every minimizer of the dual problem (see (2.7)) has a flat truncation. This will be shown in Section 2.

Second, we consider asymptotical convergence. Let  $\{y^{(k)}\}_{k=1}^{\infty}$  be a sequence of asymptotically optimal solutions of (1.4). We prove that: under the archimedean condition, for any fixed order  $t \geq \max\{d_f, d_g + |S| - 1\}$ , the truncated sequence  $\{y^{(k)}|_{2t}\}_{k=1}^{\infty}$  is bounded, and its every limit point is flat, i.e., flat truncation is asymptotically satisfied for Putinar type Lasserre's hierarchy. A similar result holds for the Schmüdgen type one when  $K$  is compact. This will be shown in Section 3.

Third, we consider the applications of the above results in unconstrained polynomial optimization and Jacobian SDP relaxations. For them, we show that flat truncation could be used as a certificate for checking finite convergence of the hierarchy of the SDP relaxations. This will be shown in Section 4.

---

<sup>1</sup>If  $f_k = f_{\min}$  and (1.1) has a minimizer, say  $x^*$ , then (1.4) always has an optimizer that is flat, e.g.,  $[x^*]_{2k}$ . Thus, it is more sensible to ask whether every optimizer has a flat truncation.

## 2. CERTIFYING FINITE CONVERGENCE

As we have seen in §1.2, flat truncation is generally a sufficient condition for Lasserre's hierarchy of (1.2) to converge in finitely many steps. In this section, we show that flat truncation is also generally a necessary condition. Thus, it could generally serve as a certificate for checking finite convergence of the hierarchy of (1.2). Lasserre's hierarchy of (1.2) is also called a Putinar type one, since it uses Putinar's Positivstellensatz [16] in representing positive polynomials. A refining of (1.2) is Schmüdgen type Lasserre's hierarchy, which uses cross products of the constraining polynomials. Similarly, flat truncation would also be generally used as a certificate for checking finite convergence of the Schmüdgen type one. We present the results in two separate subsections.

**2.1. Putinar Type Lasserre's relaxation.** The quadratic module generated by the tuple  $g$  is

$$Q(g) := \bigcup_{k=1}^{\infty} Q_k(g).$$

The archimedean condition (AC) requires that there exists  $\phi \in Q(g)$  such that the inequality  $\phi(x) \geq 0$  defines a compact set in  $x$ . Note that AC implies the feasible set  $K$  is compact. The convergence for Lasserre's hierarchy of (1.2) is based on Putinar's Positivstellensatz [16]: if a polynomial  $p$  is positive on  $K$  and AC holds, then  $p \in Q(g)$ . To certify finite convergence of Putinar type Lasserre's hierarchy, we need the following general assumption on  $f, g_1, \dots, g_m$ .

**Assumption 2.1.** *There exists  $\rho \in Q(g)$  such that for every  $J \subseteq \{1, \dots, m\}$  and*

$$\mathcal{V}_J := \{x \in \mathbb{R}^n : g_j(x) = 0 \ (\forall j \in J), \text{ rank } G_J(x) \leq |J|\},$$

*the intersection  $\mathcal{V}_J \cap \mathcal{M} \cap \mathcal{P}$  is finite. Here, denote  $J = \{j_1, \dots, j_l\}$ ,*

$$G_J(x) := [\nabla f(x) \quad \nabla g_{j_1}(x) \quad \cdots \quad \nabla g_{j_l}(x)],$$

*and  $\mathcal{M} := \{x \in \mathbb{R}^n : f(x) = f_{\min}\}$ ,  $\mathcal{P} := \{x \in \mathbb{R}^n : \rho(x) \geq 0\}$ .*

Assumption 2.1 requires that for every  $J$  the optimization problem

$$\min f(x) \quad \text{s.t.} \quad g_j(x) = 0 \ (j \in J)$$

has finitely many critical points lying on  $\mathcal{M} \cap \mathcal{P}$  (if  $u$  is a critical point of the above, then  $\text{rank } G_J(u) \leq |J|$ , cf. [13, §2]). Let  $S$  be the set of global minimizers of (1.1) and

$$\mathcal{V} := \bigcup_J \mathcal{V}_J.$$

Clearly,  $S \subseteq \mathcal{V}$ , and Assumption 2.1 implies that  $S$  is finite. This assumption is very general, as one could see from below:

- When every  $\mathcal{V}_J$  is finite, which is true for general  $f, g_1, \dots, g_m$  (cf. [13, Prop. 2.1]), Assumption 2.1 could be satisfied by choosing  $\rho = 0$ .
- Suppose  $S$  is finite,  $\mathcal{V} \cap \mathcal{M}$  is infinite, but  $(\mathcal{V} \cap \mathcal{M}) \setminus K$  belongs to a compact set  $T$  not intersecting  $K$ . Then, there exists  $\delta > 0$  such that

$$\text{dist}(x, K) \geq \delta \ \forall x \in T, \quad \text{dist}(x, K) = 0 \ \forall x \in K.$$

(Here  $\text{dist}(x, K) = \min_{u \in K} \|x - u\|_2$  and  $\|\cdot\|_2$  denotes the standard 2-norm.) The function  $\tau(x) := \text{dist}(x, K) - \delta/2$  is continuous in  $x$ . Then,

by the compactness of  $K$  and  $T$ , there exists a polynomial  $\eta$  that is an approximation of  $\tau$  and satisfies

$$\eta(x) \geq \delta/4 \quad \forall x \in T, \quad \eta(x) \leq -\delta/4 \quad \forall x \in K.$$

Clearly,  $-\eta$  is positive on  $K$ . Assume AC holds for  $g$ , then  $-\eta \in Q(g)$  by Putinar's Positivstellensatz. Assumption 2.1 could be satisfied by choosing  $\rho = -\eta$ , because  $\mathcal{V} \cap \mathcal{M} \cap \mathcal{P} = S$  is finite.

- Suppose  $\mathcal{V} \cap \mathcal{M}$  is unbounded but, except finitely many points, lies on a real algebraic variety

$$\{x \in \mathbb{R}^n : w_1(x) = \cdots = w_r(x) = 0\}$$

not intersecting  $K$ . Suppose AC holds for  $g$ , then  $K$  is compact and there exists  $\epsilon > 0$  such that the polynomial  $w := w_1^2 + \cdots + w_r^2 - \epsilon$  is positive on  $K$ . By Putinar's Positivstellensatz,  $w \in Q(g)$ . Assumption 2.1 could be satisfied by choosing  $\rho = w$ .

Our main result of this subsection is the following theorem.

**Theorem 2.2.** *Suppose Assumption 2.1 holds, the set  $S$  of global minimizers of (1.1) is nonempty, and for  $k$  big enough the optimal value of (1.2) is achievable and there is no duality gap between (1.2) and (1.4). Then, Lasserre's hierarchy of (1.2) has finite convergence if and only if every minimizer of (1.4) has a flat truncation for  $k$  sufficiently large.*

**Remark 2.3.** In Theorem 2.2, 1) if  $K$  has nonempty interior, then for every order  $k$  (1.2) achieves its optimum, and it has no duality gap; 2) if finite convergence occurs, then (1.4) always has a minimizer for  $k$  big enough (e.g.,  $[x^*]_{2k}$  is one for any  $x^* \in S$ ), and there is no duality gap; 3) when  $k$  is big enough, for every minimizer  $y^*$  of (1.4),  $y^*|_{2k-2}$  is always flat, as shown in the proof later; 4) we did not assume  $K$  is compact or the archimedean condition holds.

**Remark 2.4.** As pointed out in Laurent's survey [10, §6.6], we could get the following properties about flat truncation. 1) In Theorem 2.2, Assumption 2.1 implies  $S$  is finite. If  $\text{rank } M_k(y^*)$  is maximum over the optimizers of (1.4), then for any flat truncation of  $y^*$ , say,  $y^*|_{2t}$ ,  $\text{rank } M_t(y^*) = |S|$ . This means that all the minimizers of (1.1) could be extracted from  $y^*|_{2t}$ . 2) When  $S$  is infinite and  $\text{rank } M_k(y^*)$  is maximum over the optimizers of (1.4), then  $y^*$  could not have a flat truncation. 3) When (1.2) and (1.4) are solved by primal-dual interior-point algorithms, a minimizer  $y^*$  near the analytic center of the face of optimizers of (1.4) is usually returned and  $\text{rank } M_k(y^*)$  is typically maximum. Therefore, if the conditions of Theorem 2.2 are satisfied, by solving a sufficient high order Lasserre's relaxation via interior point methods, then we could typically find all minimizers of (1.1) if  $S$  is finite, and flat truncation could not be satisfied if  $S$  is infinite.

To prove Theorem 2.2, we need some properties about the kernels of moment and localizing matrices. Given a polynomial  $p(x)$ , denote its coefficient vector by  $p$ . For a localizing matrix  $L_h^{(k)}(y)$ , recall that if  $\deg(hp^2) \leq 2k$  then

$$p^T \cdot L_h^{(k)}(y) \cdot p = \mathcal{L}_y(hp^2).$$

If  $L_h^{(k)}(y) \cdot p = 0$ , we say  $p \in \ker L_h^{(k)}(y)$ . When  $M_k(y) = L_1^{(k)}(y)$  is a moment matrix, we similarly say  $p \in \ker M_k(y)$  if  $M_k(y) \cdot p = 0$ .

**Lemma 2.5.** *Let  $y \in \mathcal{M}_{n,2k}, h \in \mathbb{R}[x]$  be such that  $L_h^{(k)}(y) \succeq 0$ .*

- i) ([7, 10]) *Suppose  $M_k(y) \succeq 0$ . Let  $p, q \in \mathbb{R}[x]$ . If  $\deg(pq) \leq k-1$  and  $q \in \ker M_k(y)$ , then  $pq \in \ker M_k(y)$ . If  $q^\ell \in \ker M_k(y)$  and  $2\lceil \ell/2 \rceil \deg(q) \leq k-1$ , then  $q \in \ker M_k(y)$ .*
- ii) ([4]) *Let  $s$  be an SOS polynomial with  $\deg(hs) \leq 2k$ . Then,  $\langle hs, y \rangle \geq 0$ . If  $\langle hs, y \rangle = 0$ , then for any  $\phi \in \mathbb{R}[x]_{2\ell}$  with  $\deg(hs) + 2\ell \leq 2k-2$  we have  $\langle hs\phi, y \rangle = 0$ .*
- iii) *Let  $\{p_j\}_{j=1}^\infty \subset \mathbb{R}[x]$  be a sequence such that each  $\deg(hp_j^2) \leq 2k$  and*

$$L_h^{(k)}(y) \cdot p_j \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

*If  $q \in \mathbb{R}[x]$  and every  $\deg(p_j q) \leq k - \lceil \deg(h)/2 \rceil - 1$ , then*

$$\lim_{j \rightarrow \infty} L_h^{(k)}(y) \cdot (p_j q) = 0.$$

- iv) *Let  $\{s_j\}_{j=1}^\infty$  be a sequence of SOS polynomials such that*

$$\deg(hs_j) < 2k \quad (\forall j), \quad \text{and} \quad \langle hs_j, y \rangle \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

*If  $\phi \in \mathbb{R}[x]_{2\ell}$  and every  $\deg(hs_j) + 2\ell \leq 2k-2$ , then*

$$\lim_{j \rightarrow \infty} \langle hs_j \phi, y \rangle = 0.$$

*Proof.* i) The first part is basically from [10, Lemma 5.7] or [9, Lemma 21]. For the second part, if  $\ell$  is even, the result follows Lemma 3.9 of [7]; if  $\ell$  is odd, then  $q^{\ell+1} \in \ker M_k(y)$  from the first part, and the result is still true.

ii) By linearity of the Riesz functional  $\mathcal{L}_y(\cdot) = \langle \cdot, y \rangle$ , we would generally assume  $s = p^2$  is a single square. Then

$$\langle hs, y \rangle = \langle hp^2, y \rangle = p^T L_h^{(k)}(y) p \geq 0.$$

Now, assume  $\langle hs, y \rangle = 0$ . One could always write  $\phi = \sum_i q_i^2 - \sum_j p_j^2$  with all  $q_i, p_j \in \mathbb{R}[x]_\ell$ . By the linearity of  $\langle \cdot, y \rangle$  again, we could further assume  $\phi = q^2$  is a single square. Note that

$$\langle hp^2, y \rangle = p^T L_h^{(k)}(y) p, \quad \langle hp^2 q^2, y \rangle = (pq)^T L_h^{(k)}(y) (pq).$$

Since  $L_h^{(k)}(y) \succeq 0$ ,  $\langle hs, y \rangle = 0$  implies  $L_h^{(k)}(y) p = 0$ . By Lemma 2.2 of [4], we also have  $L_h^{(k)}(y) (pq) = 0$ . Then  $(pq)^T L_h^{(k)}(y) (pq) = 0$ , i.e.,  $\langle hs\phi, y \rangle = 0$ .

iii) Let  $r = k - \lceil \deg(h)/2 \rceil$ . By a simple induction on  $\deg(q)$ , it suffices to prove the lemma for  $q = x_i$  ( $1 \leq i \leq n$ ). Note that every  $\deg(p_j) + 1 \leq r-1$ . Let

$$u^{(j)} := L_h^{(k)}(y) \cdot p_j, \quad v^{(j)} := L_h^{(k)}(y) \cdot (p_j q).$$

They are indexed by integral vectors  $\alpha \in \mathbb{N}^n$ , and could also be expressed as

$$u^{(j)} = \mathcal{L}_y(p_j \cdot h \cdot [x]_r), \quad v^{(j)} = \mathcal{L}_y(x_i \cdot p_j \cdot h \cdot [x]_r).$$

Then  $u^{(j)} \rightarrow 0$  implies

$$\lim_{j \rightarrow \infty} \mathcal{L}_y(p_j \cdot h \cdot x^\beta) = 0 \quad \text{if } |\beta| \leq r.$$

The  $\alpha$ -th entry of  $v^{(j)}$  is

$$v_\alpha^{(j)} := \mathcal{L}_y(p_j \cdot h \cdot x^\alpha \cdot x_i).$$



If  $\deg(x^\alpha \cdot x_i) \leq r$ , then  $v_\alpha^{(j)} = u_{\tilde{\alpha}}^{(j)} \rightarrow 0$  ( $\tilde{\alpha}$  is the exponent of  $x^\alpha \cdot x_i$ ). So,

$$\lim_{j \rightarrow \infty} v_\alpha^{(j)} = 0 \quad \text{if } |\alpha| \leq r - 1.$$

This implies (note  $L_h^{(k-1)}(y)$  is a leading principal submatrix of  $L_h^{(k)}(y)$ ) that

$$\lim_{j \rightarrow \infty} \left( L_h^{(k-1)}(y) \right) \cdot (p_j \cdot x_i) = \lim_{j \rightarrow \infty} \mathcal{L}_y(x_i \cdot p_j \cdot h \cdot [x]_{r-1}) = 0.$$

Since  $L_h^{(k-1)}(y)$  is symmetric, from the above, we know

$$\lim_{j \rightarrow \infty} (p_j \cdot x_i)^T \cdot \left( L_h^{(k-1)}(y) \right) \cdot (p_j \cdot x_i) = 0.$$

When  $|\alpha| > r - 1$ , the coefficient of  $x^\alpha$  in the polynomial  $p_j \cdot x_i$  is zero. So

$$(p_j \cdot x_i)^T \cdot \left( L_h^{(k)}(y) \right) \cdot (p_j \cdot x_i) = (p_j \cdot x_i)^T \cdot \left( L_h^{(k-1)}(y) \right) \cdot (p_j \cdot x_i).$$

Hence, we also have

$$\lim_{j \rightarrow \infty} (p_j \cdot x_i)^T \cdot \left( L_h^{(k)}(y) \right) \cdot (p_j \cdot x_i) = 0.$$

Since  $L_h^{(k)}(y) \succeq 0$ , the above implies that

$$\lim_{j \rightarrow \infty} \left( L_h^{(k)}(y) \right) \cdot (p_j \cdot x_i) = 0.$$

iv) Write  $s_j = \sum_{\ell} p_{j,\ell}^2$  (the length of summation is at most  $\binom{n+k}{k}$ , since  $\deg(s_j) \leq 2k$ ). By part ii),  $\langle h s_j, y \rangle \geq \langle h p_{j,\ell}^2, y \rangle \geq 0$ . So,  $\langle h s_j, y \rangle \rightarrow 0$  implies

$$\langle h p_{j,\ell}^2, y \rangle = (p_{j,\ell})^T L_h^{(k)}(y) p_{j,\ell} \rightarrow 0.$$

Since  $L_h^{(k)}(y) \succeq 0$ , we have  $L_h^{(k)}(y) p_{j,\ell} \rightarrow 0$  as  $j \rightarrow \infty$ . Like in item ii), we would also assume  $\phi = q^2$  is a single square. By part iii), it holds that

$$L_h^{(k)}(y) (p_{j,\ell} q) \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

This implies that  $(p_{j,\ell} q)^T L_h^{(k)}(y) (p_{j,\ell} q) \rightarrow 0$  and

$$\langle h s_j \phi, y \rangle = \sum_{\ell} \langle h p_{j,\ell}^2 q^2, y \rangle = \sum_{\ell} (p_{j,\ell} q)^T L_h^{(k)}(y) (p_{j,\ell} q) \rightarrow 0.$$

□

*Proof of Theorem 2.2* The sufficiency of flat truncation was observed in §1.2, if there is no duality gap between (1.2) and (1.4). We only need to prove its necessity. Suppose  $f_{k_0} = f_{\min}$  for  $k_0$  big enough. Since (1.2) has a maximizer when its order is big enough, we could assume (1.2) achieves its optimum  $f_{\min}$  for order  $k_0$  (otherwise increase  $k_0$ ), i.e.,  $f(x) - f_{\min} \in Q_{k_0}(g)$ . So, there exist SOS polynomials  $s_0, s_1, \dots, s_m$  such that every  $\deg(g_i s_i) \leq 2k_0$  and

$$(2.1) \quad f - f_{\min} = s_0 + g_1 s_1 + \dots + g_m s_m.$$

Let  $y^*$  be an arbitrary minimizer of (1.4) (it always has one when  $k \geq k_0$ , e.g.,  $[x^*]_{2k}$ , for any  $x^* \in S$ ). Clearly,  $\langle f, y^* \rangle = f_{\min}$ . Let  $C$  be the semialgebraic set defined as ( $\rho$  is from Assumption 2.1)

$$(2.2) \quad C = \{x \in \mathbb{R}^n : s_0(x) = g_1(x) s_1(x) = \dots = g_m(x) s_m(x) = 0, \rho(x) \geq 0\}.$$

We complete the proof in three steps.



**Step 1** We show that  $C$  is a finite set. In the identity (2.1), differentiating its both sides in  $x$  results in (note  $g_0 \equiv 1$ )

$$\nabla f(x) = \sum_{i=0}^m \left( s_i(x) \cdot \nabla g_i(x) + g_i(x) \nabla s_i(x) \right).$$

Choose an arbitrary  $u \in C$ , then clearly  $u \in \mathcal{M} \cap \mathcal{P}$ . Let  $J(u) = \{i : g_i(u) = 0\}$ . Note that for every  $i \notin J(u)$ ,  $s_i(u) = 0$ , and it implies  $\nabla s_i(u) = 0$  (because  $s_i$  is SOS and  $u$  is a minimizer of  $s_i$ ). So, by the above, it holds that

$$\nabla f(u) = \sum_{i \in J(u)} s_i(u) \cdot \nabla g_i(u).$$

Hence,  $u \in \mathcal{V}_{J(u)} \cap \mathcal{M} \cap \mathcal{P}$ . Since there are at most  $2^m$  active index sets like  $J(u)$ , by Assumption 2.1,  $C$  must be finite.

**Step 2** We show that every generator of the vanishing ideal of  $C$  belongs to the kernel of  $M_k(y^*)$  for  $k$  big enough. Since  $C$  is finite, its vanishing ideal

$$I(C) := \{p \in \mathbb{R}[x] : p(u) = 0 \quad \forall u \in C\}$$

is zero dimensional. Let  $\{h_1, \dots, h_r\}$  be a Grobner basis of  $I(C)$  with respect to a total degree ordering. Clearly, each  $h_i$  vanishes on  $C$ . By Positivstellensatz (cf. Corollary 4.4.3 of [1]), there exist  $\ell \in \mathbb{N}$ ,  $\phi_1, \dots, \phi_m \in \mathbb{R}[x]$ , and  $\varphi \in Q(\rho)$  (the quadratic module generated by the single polynomial  $\rho$ , which is also equal to the preordering generated by  $\rho$ ) such that

$$h_i^{2\ell} + \varphi + g_1 s_1 \phi_1 + \dots + g_m s_m \phi_m = 0.$$

Applying the Riesz functional  $\mathcal{L}_{y^*}$  to the above (suppose  $2k$  is bigger than the degrees of all the products there), we get

$$(2.3) \quad \langle h_i^{2\ell}, y^* \rangle + \langle \varphi, y^* \rangle + \sum_{j=1}^m \langle g_j s_j \phi_j, y^* \rangle = 0.$$

Applying  $\mathcal{L}_{y^*}$  to (2.1) results in (note  $\langle f, y^* \rangle = f_{\min}$ )

$$(2.4) \quad 0 = \langle f - f_{\min}, y^* \rangle = \sum_{j=0}^m \langle g_j s_j, y^* \rangle.$$

Since each  $s_j$  is SOS, every  $\langle g_j s_j, y^* \rangle \geq 0$ , by item ii) of Lemma 2.5. Thus, from the above, we know every  $\langle g_j s_j, y^* \rangle = 0$ . Again, by item ii) of Lemma 2.5, if  $2k > 2 + \deg(g_j s_j \phi_j)$ , then every  $\langle g_j s_j \phi_j, y^* \rangle = 0$ . So, from (2.3), we can get

$$(2.5) \quad \langle h_i^{2\ell}, y^* \rangle + \langle \varphi, y^* \rangle = 0.$$

Since  $h_i^{2\ell}$  is SOS, we similarly have  $\langle h_i^{2\ell}, y^* \rangle \geq 0$ . Since  $Q(\rho) \subset Q(g)$ ,  $\varphi \in Q(g)$  and one could write  $\varphi = \sum_{j=0}^m g_j \sigma_j$  with each  $\sigma_j$  being SOS. Hence,

$$\langle \varphi, y^* \rangle = \langle g_0 \sigma_0, y^* \rangle + \langle g_1 \sigma_1, y^* \rangle + \langle g_m \sigma_m, y^* \rangle \geq 0,$$

by item ii) of Lemma 2.5. Then (2.5) implies  $\langle h_i^{2\ell}, y^* \rangle = 0$ , i.e.,  $h_i^\ell \in \ker M_k(y^*)$ . By item i) of Lemma 2.5 and  $M_k(y^*) = L_{g_0}^{(k)}(y^*) \succeq 0$ , if  $k$  is big enough, we get

$$h_i \in \ker M_k(y^*).$$

**Step 3** It's enough to show that  $y^*|_{2k-2}$  is flat. Since  $C$  is finite, the quotient space  $\mathbb{R}[x]/I(C)$  is finitely dimensional. Let  $\{b_1, \dots, b_L\}$  be a standard basis of  $\mathbb{R}[x]/I(C)$ . Then, for every  $\alpha \in \mathbb{N}^n$ , we can write

$$x^\alpha = \eta(\alpha) + \sum_{i=1}^r \theta_i h_i, \quad \deg(\theta_i h_i) \leq |\alpha|, \quad \eta(\alpha) \in \text{span}\{b_1, \dots, b_L\}.$$

Because every  $h_i \in \ker M_k(y^*)$ , we know

$$\theta_i h_i \in \ker M_k(y^*) \quad \text{if } |\alpha| \leq k-1,$$

by applying item i) of Lemma 2.5. Thus,

$$x^\alpha - \eta(\alpha) \in \ker M_k(y^*) \quad \text{if } |\alpha| \leq k-1.$$

Set  $d_b := \max_j \deg(b_j)$ . Then every  $\alpha$ -th column ( $d_b + 1 \leq |\alpha| \leq k-1$ ) of  $M_k(y^*)$  is a linear combination of  $\beta$ -th columns of  $M_k(y^*)$  with  $|\beta| \leq d_b$ , so

$$\text{rank } M_{d_b}(y^*) = \text{rank } M_t(y^*), \quad t = d_b + 1, \dots, k-1.$$

Hence, if  $k-1-d_g \geq d_b$ , then

$$\text{rank } M_{k-1-d_g}(y^*) = \text{rank } M_{k-1}(y^*).$$

That is,  $y^*|_{2k-2}$  is flat and  $y^*$  has a flat truncation, when  $k$  is big enough.  $\square$

**2.2. Schmüdgen type Lasserre's relaxation.** Now we consider Schmüdgen type Lasserre's hierarchy, which refines (1.2) as:

$$(2.6) \quad \max \quad \gamma \quad \text{s.t.} \quad f(x) - \gamma \in Pr_k(g).$$

The above  $Pr_k(g)$  denotes the  $k$ -th truncated preordering generated by the tuple  $g$  (denote  $g_\nu := g_1^{\nu_1} \cdots g_m^{\nu_m}$ ):

$$Pr_k(g) := \left\{ \sum_{\nu \in \{0,1\}^m} g_\nu \sigma_\nu \mid \sigma_\nu \text{ is SOS, } \deg(g_\nu \sigma_\nu) \leq 2k \text{ for every } \nu \right\}.$$

The dual optimization problem of (2.6) is (cf. [6, 8])

$$(2.7) \quad \begin{cases} \min & \langle f, y \rangle \\ \text{s.t.} & L_{g_\nu}^{(k)}(y) \succeq 0 \ (\nu \in \{0,1\}^m), \langle 1, y \rangle = 1. \end{cases}$$

Let  $\tilde{f}_k$  and  $\tilde{f}_k^*$  be the optimal values of (2.6) and (2.7) respectively, for a given order  $k$ . By weak duality,  $\tilde{f}_k \leq \tilde{f}_k^*$  for every  $k$ . If  $K$  has nonempty interior, then  $\tilde{f}_k = \tilde{f}_k^*$ , i.e., there is no duality gap. Clearly, every  $\tilde{f}_k^* \leq f_{\min}$ . Both sequences  $\{\tilde{f}_k\}$  and  $\{\tilde{f}_k^*\}$  are monotonically increasing. The relaxation (2.6) is stronger than (1.2), because  $Q_k(g) \subseteq Pr_k(g)$ , so we have  $f_k \leq \tilde{f}_k$  for every  $k$ . By Schmüdgen's Positivstellensatz (if  $K$  is compact and a polynomial  $p$  is positive on  $K$ , then  $p \in Pr_\ell(g)$  for some  $\ell$ , cf. [18]), then  $\{\tilde{f}_k\}$  converges to  $f_{\min}$ , and so does  $\{\tilde{f}_k^*\}$ , when  $K$  is compact. If  $\tilde{f}_{k_1} = f_{\min}$  for some order  $k_1$ , we say Schmüdgen type Lasserre's hierarchy has *finite convergence*.

Suppose  $y^*$  is an optimizer of (2.7). If  $y^*$  has a flat truncation, say  $y^*|_{2t}$ , then, as shown in §1.2, one could not only extract  $\text{rank } M_t(y^*)$  global optimizers of (1.1) from  $y^*$ , but also get a certificate for  $\tilde{f}_k = f_{\min}$  if there is no duality gap between (2.6) and (2.7). So, flat truncation is also generally a sufficient condition for Schmüdgen type Lasserre's hierarchy to have finite convergence. Is it also generally a necessary

condition? If so, flat truncation would also serve as a certificate for checking finite convergence of the hierarchy of (2.6). Like for the Putinar type one, this is also generally true. A similar result like Theorem 2.2 holds, and weaker conditions are required.

**Theorem 2.6.** *Suppose the set  $S$  of global minimizers of (1.1) is nonempty and finite, and there is no duality gap between (2.6) and (2.7) for  $k$  big enough. Then, Schmüdgen type Lasserre's hierarchy of (2.6) has finite convergence if and only if every minimizer of (2.7) has a flat truncation for  $k$  sufficiently large.*

**Remark 2.7.** The comments in Remarks 2.3 and 2.4 for Theorem 2.2 are all applicable to Theorem 2.6. Here, we point out some differences: 1) Theorem 2.6 does not require the optimum of (2.6) to be achievable. 2) Assumption 2.1 is slightly stronger than that  $S$  is finite, although they are both generally true. 3) If it occurs that  $Q_k(g) = Pr_k(g)$  (e.g., this is the case when (1.1) has equality constraints and/or a single inequality constraint), then in Theorem 2.2 Assumption 2.1 could be replaced by  $|S| < \infty$  and (1.2) is not required to achieve its optimum.

*Proof of Theorem 2.6* Like in the proof of Theorem 2.2, we only need to prove the necessity of flat truncation. Suppose  $\tilde{f}_{k_1} = f_{min}$  for some order  $k_1$ , then for every  $\epsilon > 0$  we have  $f - f_{min} + \epsilon \in Pr_{k_1}(g)$ . Write

$$(2.8) \quad f - f_{min} + \epsilon = \sum_{\nu \in \{0,1\}^m} \sigma_\nu^\epsilon \cdot g_\nu$$

for some SOS polynomials  $\sigma_\nu^\epsilon$  with  $\deg(\sigma_\nu^\epsilon g_\nu) \leq 2k_1$ . Note that as  $\epsilon \rightarrow 0$  some coefficients of  $\sigma_\nu^\epsilon$  might go to infinity while its degree is bounded. Let  $y^*$  be an arbitrary minimizer of (2.7) (if  $k \geq k_1$ , (2.7) always has one, e.g.,  $[x^*]_{2k}$ , for any  $x^* \in S$ ). Then  $\mathcal{L}_{y^*}(f) = \tilde{f}_k = f_{min}$ . Applying  $\mathcal{L}_{y^*}$  to (2.8), we get

$$(2.9) \quad \epsilon = \sum_{\nu \in \{0,1\}^m} \langle \sigma_\nu^\epsilon g_\nu, y^* \rangle.$$

Since every  $\sigma_\nu^\epsilon$  is SOS, by item ii) of Lemma 2.5,  $\langle \sigma_\nu^\epsilon g_\nu, y^* \rangle \geq 0$  and

$$(2.10) \quad \lim_{\epsilon \rightarrow 0} \langle \sigma_\nu^\epsilon g_\nu, y^* \rangle = 0.$$

We complete the proof in three steps.

**Step 1** The set  $S$  is finite, so its vanishing ideal

$$I(S) = \{c \in \mathbb{R}[x] : c(u) = 0 \quad \forall u \in S\}$$

is zero dimensional. Let  $\{h_1, \dots, h_r\}$  be a Grobner basis of  $I(S)$  with respect to a total degree ordering. Clearly, each  $h_i$  vanishes on

$$S = \{x \in \mathbb{R}^n : -(f(x) - f_{min}) = 0, g_1(x) \geq 0, \dots, g_m(x) \geq 0\}.$$

By Positivstellensatz (cf. Corollary 4.4.3 of [1]), there exist  $\ell \in \mathbb{N}$ ,  $\varphi \in \mathbb{R}[x]$  and SOS polynomials  $\phi_\nu$  ( $\nu \in \{0,1\}^m$ ) such that

$$h_i^{2\ell} - (f - f_{min})\varphi + \sum_{\nu \in \{0,1\}^m} \phi_\nu g_\nu = 0.$$

Applying  $\mathcal{L}_{y^*}$  to the above (suppose  $2k$  is bigger than the degrees of all the above products) results in

$$(2.11) \quad \langle h_i^{2\ell}, y^* \rangle + \sum_{\nu \in \{0,1\}^m} \langle \phi_\nu g_\nu, y^* \rangle = \langle (f - f_{\min})\varphi, y^* \rangle.$$

By (2.8), for every  $\epsilon > 0$ , we get

$$(2.12) \quad \langle (f - f_{\min} + \epsilon)\varphi, y^* \rangle = \sum_{\nu \in \{0,1\}^m} \langle g_\nu \sigma_\nu^\epsilon \varphi, y^* \rangle.$$

**Step 2** By (2.10) and item iv) of Lemma 2.5, we can get

$$\lim_{\epsilon \rightarrow 0} \langle g_\nu \sigma_\nu^\epsilon \varphi, y^* \rangle = 0$$

for  $k$  big enough. Hence, from (2.12) and the above, it holds that

$$\langle (f - f_{\min})\varphi, y^* \rangle = \lim_{\epsilon \rightarrow 0} \langle (f - f_{\min} + \epsilon)\varphi, y^* \rangle = 0.$$

So, (2.11) results in the equality

$$\langle h_i^{2\ell}, y^* \rangle + \sum_{\nu \in \{0,1\}^m} \langle \phi_\nu g_\nu, y^* \rangle = 0.$$

Since  $h_i^{2\ell}$  and every  $\phi_\nu$  are SOS, each  $\langle \phi_\nu g_\nu, y^* \rangle \geq 0$  and  $\langle h_i^{2\ell}, y^* \rangle \geq 0$ , by item ii) of Lemma 2.5. Hence,  $\langle h_i^{2\ell}, y^* \rangle = 0$ , i.e.,  $h_i^\ell \in \ker M_k(y^*)$ . Again, by item i) of Lemma 2.5 and  $M_k(y^*) = L_{g_0}^{(k)}(y^*) \succeq 0$ , if  $k$  is big enough, then

$$h_i \in \ker M_k(y^*).$$

**Step 3** Like Step 3 in the proof of Theorem 2.2, we would prove  $y^*|_{2k-2}$  is flat by repeating the same argument there, and omit it here for cleanness.  $\square$

### 3. ASYMPTOTICAL CONVERGENCE

In this section, we consider the general case that Lasserre's hierarchy of (1.2) has asymptotic but not finite convergence. Under the archimedean condition, Lasserre proved  $f_k \rightarrow f_{\min}$  as  $k \rightarrow \infty$ . Since it is possible that  $f_k = f_k^* < f_{\min}$  for every  $k$ , we should not expect flat truncation holds in such cases. When (1.1) has a unique global minimizer  $u^*$ , Schweighofer [20] showed an important result: the subvector consisting of linear moments (indexed by  $\alpha \in \mathbb{N}^n$  with  $|\alpha| = 1$ ) of an almost optimizer  $y^{(k)}$  of (1.4) for order  $k$  converges to  $u^*$  as  $k \rightarrow \infty$ . When (1.1) has more than one minimizer, do we have a similar convergence result? To the author's best knowledge, this question was known very little. This section is addressing this issue. Generally, we need to use a higher order truncation of  $y^{(k)}$ , and consider its limit points. The main result of this section is that every limit point of a truncation of  $y^{(k)}$  is flat if (1.1) has finitely many global minimizers. In other words, flat truncation is asymptotically satisfied for the general case.

We assume the archimedean condition holds for the tuple  $g$ : there exist  $R > 0$  and  $k_0 \in \mathbb{N}$  such that

$$(3.1) \quad R - \|x\|_2^2 \in Q_{k_0}(g).$$

For a Borel set  $T \subseteq \mathbb{R}^n$ , denote by  $\mathbf{Prob}(T)$  the set of all probability measures supported on  $T$ . Let  $S$  be the set of global minimizers of (1.1). For each integer  $t > 0$ , denote

$$(3.2) \quad F_{2t} := \left\{ \int [x]_{2t} d\mu : \mu \in \mathbf{Prob}(S) \right\}.$$

**Proposition 3.1.** *Let  $S$  and  $F_{2t}$  be defined as above. If  $0 < |S| < \infty$ , then for any integer  $t \geq \max\{d_f, d_g + |S| - 1\}$ , every  $z \in F_{2t}$  is flat.*

*Proof.* Let  $\ell = |S|$  and write  $S = \{(a_{i,1}, \dots, a_{i,n}) : i = 1, \dots, \ell\}$ , and  $I$  be the ideal generated by the following polynomial equations

$$(x_{i_1} - a_{1,i_1})(x_{i_2} - a_{2,i_2}) \cdots (x_{i_\ell} - a_{\ell,i_\ell}) = 0, \quad \forall i_1, \dots, i_\ell \in \{1, \dots, n\}.$$

Clearly, the zero set of the above equations is  $S$ . For each  $\alpha \in \mathbb{N}^n$  with  $|\alpha| = \ell$ , the  $x^\alpha$  is a leading monomial in one of the above defining polynomials, so there exists  $p_\alpha \in \mathbb{R}[x]_{\ell-1}$  such that

$$(3.3) \quad x^\alpha - p_\alpha(x) \equiv 0 \pmod{I}.$$

Choose an arbitrary  $z \in F_{2t}$ , then there exists  $\mu \in \mathbf{Prob}(S)$  satisfying

$$z = \int_S [x]_{2t} d\mu, \quad M_t(z) = \int_S [x]_t [x]_t^T d\mu.$$

By a simple induction on  $|\alpha|$ , one could show that for every  $|\alpha| \in [\ell, t]$ , there exists  $p_\alpha \in \mathbb{R}[x]_{\ell-1}$  satisfying (3.3). Clearly,  $x^\alpha - p_\alpha(x)$  vanishes on  $S$  and

$$x^\alpha - p_\alpha(x) \in \ker M_t(z).$$

This means that every  $\alpha$ -th ( $|\alpha| \geq \ell$ ) column of  $M_t(z)$  is a linear combination of its  $\beta$ -th columns ( $|\beta| \leq \ell - 1$ ). So, it holds that

$$\text{rank } M_{\ell-1}(z) = \text{rank } M_r(z), \quad \forall r = \ell, \ell + 1, \dots, t.$$

Thus, for all  $t \geq d_g + \ell - 1$ , we have

$$\text{rank } M_{t-d_g}(z) = \text{rank } M_t(z).$$

Clearly, every  $L_{g_i}^{(t)}(z) \succeq 0$ , and hence  $z$  is flat by the definition.  $\square$

Denote by  $\mathcal{M}_\infty$  the space of all full moment sequences  $w = (w_\alpha)$  indexed by vectors  $\alpha \in \mathbb{N}^n$ . For all  $w, z \in \mathcal{M}_\infty$ , define

$$\langle w, z \rangle = \sum_{\alpha} w_{\alpha} z_{\alpha}, \quad \|w\|_2 = \sqrt{\langle w, w \rangle}.$$

This induces a Hilbert space

$$\mathcal{M}_\infty^2 = \{w \in \mathcal{M}_\infty : \|w\|_2 < \infty\}.$$

If we think of  $\mathcal{M}_\infty^2$  as a Banach space, then it is self-dual. Clearly, (1.1) is equivalent to the optimization problem:

$$(3.4) \quad \min_{y \in \mathcal{M}_\infty} \langle f, y \rangle \quad \text{s.t.} \quad y \in \left\{ \int_K [x]_\infty d\mu : \mu \in \mathbf{Prob}(K) \right\}.$$

The set of its optimizers is precisely  $\mathbf{Prob}(S)$ .

**Lemma 3.2.** *Suppose (3.1) holds. If  $y$  is feasible for (1.4) and  $k \geq k_1 \geq k_0$ , then*

$$(3.5) \quad \|y|_{2(k_1-k_0)}\|_2^2 \leq 1 + R + \dots + R^{k_1-k_0}.$$

*Proof.* By (3.1), there exist SOS polynomials  $s_0, s_1, \dots, s_m$  such that

$$R - \|x\|_2^2 = \sum_{i=0}^m g_i s_i, \quad \text{each } \deg(g_i s_i) \leq 2k_0.$$

(Here  $\|x\|_2 := \sqrt{x_1^2 + \dots + x_n^2}$ .) Thus, for every  $j = 1, \dots, k_1 - k_0$ , we have

$$R \cdot \|x\|_2^{2j-2} - \|x\|_2^{2j} = \sum_{i=0}^m g_i \cdot s_i \cdot \|x\|_2^{2j-2}.$$

Applying  $\mathcal{L}_y$  to the above gives (by item ii) of Lemma 2.5)

$$R \mathcal{L}_y(\|x\|_2^{2j-2}) - \mathcal{L}_y(\|x\|_2^{2j}) \geq 0, \quad j = 1, \dots, k_1 - k_0.$$

Therefore, it holds that

$$\mathcal{L}_y(\|x\|_2^2) \leq R, \quad \mathcal{L}_y(\|x\|_2^4) \leq R^2, \dots, \quad \mathcal{L}_y(\|x\|_2^{2j}) \leq R^j.$$

The moment matrix  $M_k(y) = L_{g_0}^{(k)}(y) \succeq 0$  implies its submatrix  $M_{k_1-k_0}(y) \succeq 0$ . In the below, we denote by  $\|A\|_F$  the Frobenius norm of a matrix  $A$ , i.e.,  $\|A\|_F = \sqrt{\text{Trace}(A^T A)}$ . Recall that  $\|A\|_F \leq \text{Trace}(A)$  if  $A \succeq 0$ . Clearly,

$$\sum_{|\alpha| \leq 2(k_1-k_0)} |y_\alpha|^2 \leq \|M_{k_1-k_0}(y)\|_F^2 \leq (\text{Trace}(M_{k_1-k_0}(y)))^2,$$

$$\text{Trace}(M_{k_1-k_0}(y)) = \sum_{j=0}^{k_1-k_0} \sum_{|\alpha|=2j} y_\alpha \leq \sum_{j=0}^{k_1-k_0} \mathcal{L}_y(\|x\|_2^{2j}) \leq \sum_{j=0}^{k_1-k_0} R^j.$$

So, the inequality (3.5) is true.  $\square$

It is possible that (1.4) might not have a minimizer, i.e., the optimal value  $f_k^*$  of (1.4) may not be achievable (cf. [20, Example 4.8]). Thus, we consider an almost optimizer of (1.4). Let  $\{y^{(k)}\}$  be a sequence such that  $y^{(k)}$  is feasible for (1.4) with order  $k$ . We say  $\{y^{(k)}\}$  is *asymptotically optimal* if

$$\lim_{k \rightarrow \infty} \langle f, y^{(k)} \rangle = \lim_{k \rightarrow \infty} f_k^*.$$

(A different notion *nearly optimality* was used in [20].) Note that if (3.1) holds and  $\{y^{(k)}\}$  is asymptotically optimal, then  $\langle f, y^{(k)} \rangle \rightarrow f_{\min}$  as  $k \rightarrow \infty$ .

**Theorem 3.3.** *Assume the archimedean condition (3.1) holds and the set  $S$  of global minimizers of (1.1) is nonempty and finite. Let  $\{y^{(k)}\}$  be asymptotically optimal for (1.4). Then, for every  $t \geq \max\{d_f, d_g + |S| - 1\}$ , the truncated sequence  $\{y^{(k)}|_{2t}\}$  is bounded, and its every limit point belongs to  $F_{2t}$  and is flat.*

*Proof.* If we replace  $k_1 - k_0$  by  $t$  in (3.5) of Lemma 3.2, then the sequence  $\{y^{(k)}|_{2t}\}$  is clearly bounded. We could generally assume  $R < 1$ , because otherwise one can scale  $x$  in (3.1) so that  $R < 1$  and Lasserre's hierarchy remains equivalent.

Let  $v$  be an arbitrary limit point of  $\{y^{(k)}|_{2t}\}$ . One could generally assume  $y^{(k)}|_{2t} \rightarrow v$ . We need to prove  $v \in F_{2t}$ . Here we exploit a technique that is used in [4]. Suppose otherwise  $v \notin F_{2t}$ . For every  $k$ , define

$$z^{(k)} := y^{(k)}|_{2(k-k_0)}.$$

Each  $z^{(k)}$  could be treated as a vector in  $\mathcal{M}_\infty^2$  by adding zero entries to the tailing. So, for  $k > k_0$ , the inequality (3.5) implies

$$\|z^{(k)}\|_2^2 \leq 1 + R + R^2 + R^3 + \cdots = 1/(1 - R).$$

The sequence  $\{z^{(k)}\}$  is bounded in  $\mathcal{M}_\infty^2$ . By Alaoglu's Theorem (cf. [2, Theorem V.3.1] or [8, Theorem C.18]), it has a subsequence  $\{z^{(k_j)}\}$  that is convergent in the weak-\* topology. That is, there exists  $z^* \in \mathcal{M}_\infty^2$  such that

$$\langle c, z^{(k_j)} \rangle \rightarrow \langle c, z^* \rangle \quad \text{as } j \rightarrow \infty$$

for every  $c \in \mathcal{M}_\infty^2$ . If we choose  $c$  as  $\langle c, w \rangle = w_\alpha$  for each  $\alpha$ , then one could get

$$(3.6) \quad y^{(k_j)}|_{2t} = z^{(k_j)}|_{2t} \rightarrow z^*|_{2t} = v.$$

Note that every  $L_{g_i}^{(r)}(z^{(k_j)}) \succeq 0$  if  $k_j \geq 2r$ . By (3.6), it holds that for all  $r$

$$L_{g_i}^{(r)}(z^*) \succeq 0, \quad i = 0, 1, \dots, m.$$

Hence,  $z^* \in \mathcal{M}_\infty^2$  is a full moment sequence such that the corresponding localizing matrices of all orders are positive semidefinite. By Lemma 3.2 of Putinar [16] and (3.1),  $z^*$  admits a probability measure (note  $\mathcal{L}_{z^*}(1) = 1$ ) supported on  $K$ , and it is feasible for (3.4). The polynomial  $f$  defines a continuous linear functional acting on  $\mathcal{M}_\infty^2$  as  $\langle f, w \rangle$ . By the weak-\* convergence of  $z^{(k_j)} \rightarrow z^*$ , it holds that

$$\langle f, z^* \rangle = \lim_{j \rightarrow \infty} \langle f, z^{(k_j)} \rangle = \lim_{j \rightarrow \infty} \langle f, y^{(k_j)} \rangle = \lim_{j \rightarrow \infty} f_{k_j}^* = f_{\min}.$$

The above last two equalities are because (3.1) holds and the sequence  $\{y^{(k)}\}$  is asymptotically optimal. This means that  $z^*$  is also a minimizer of (3.4). Hence,  $v = z^*|_{2t} \in F_{2t}$ . But, this contradicts the earlier assertion that  $v \notin F_{2t}$ . Therefore, every limit point of  $\{y^{(k)}|_{2t}\}$  belongs to  $F_{2t}$ . By Proposition 3.1, we know every  $v \in F_{2t}$  is flat if  $t \geq \max\{d_f, d_g + |S| - 1\}$ .  $\square$

In Theorem 3.3, for a fixed  $t$ , it is possible that the truncation  $y^{(k)}|_{2t}$  is not flat for every  $k$ . But, for every  $\epsilon > 0$ , if  $k$  is big enough, there exists a  $z \in F_{2t}$  such that  $\|y^{(k)}|_{2t} - z\|_2 \leq \epsilon$ . In other words, flat truncation is asymptotically satisfied when  $S$  is finite. Theorem 3.3 also implies that the distance between  $y^{(k)}|_{2t}$  and  $F_{2t}$  tends to zero as  $k \rightarrow \infty$ . Therefore, in numerical computations, flat truncation has a good chance to be satisfied.

**Remark 3.4.** A similar version of Theorem 3.3 also holds for Schmüdgen type Lasserre's hierarchy. This is because (2.7) can be thought of as a Putinar type one applied to the  $2^m$  constraints  $g_\nu(x) \geq 0$  ( $\nu \in \{0, 1\}^m$ ). The archimedean condition (3.1) for the corresponding constraints is automatically satisfied for compact  $K$ , by Schmüdgen's Positivstellensatz. Hence, when  $K$  is compact and  $S$  is finite, flat truncation is also asymptotically satisfied for the hierarchy of (2.7).

#### 4. SOME APPLICATIONS

In this section, we consider two interesting cases of semidefinite programming relaxations for solving polynomial optimization, for which flat truncation would serve as a certificate to check finite convergence and be used to get minimizers.



**4.1. Unconstrained polynomial optimization.** Consider the unconstrained polynomial optimization

$$(4.1) \quad \min_{x \in \mathbb{R}^n} f(x).$$

To have a finite minimum  $f_{\min}$ , assume  $f(x)$  has an even degree  $2d$ . The standard SOS relaxation (cf. [6, 15]) for solving (4.1) is

$$(4.2) \quad \max \quad \gamma \quad \text{s.t.} \quad f(x) - \gamma \quad \text{is SOS.}$$

Its dual optimization problem is

$$(4.3) \quad \min_{y \in \mathcal{M}_{2d}} \langle f, y \rangle \quad \text{s.t.} \quad M_d(y) \succeq 0, \langle 1, y \rangle = 1.$$

Lasserre [6] showed that (4.2) and (4.3) have the same optimal value (there is no duality gap), which we denote by  $f_{\text{sos}}$ , and (4.2) achieves its optimum if  $f_{\text{sos}} > -\infty$ , because (4.3) has an interior point. Clearly,  $f_{\text{sos}} \leq f_{\min}$ . As demonstrated by the numerical experiments of [15], it occurs quite a lot that  $f_{\min} = f_{\text{sos}}$ . Thus, one is wondering: how do we check  $f_{\text{sos}} = f_{\min}$ , and if so how do we get minimizers of (4.1)? This issue could be solved by using flat truncation.

The whole space  $\mathbb{R}^n$  would be defined by the trivial inequality  $1 \geq 0$ . Thus, the associated  $k$ -th truncated quadratic module and preordering coincide and

$$Q_k(1) = Pr_k(1) = \Sigma_{n,2k},$$

where  $\Sigma_{n,2k}$  denotes the cone of SOS polynomials having  $n$  variables and degree  $2k$ . Like (1.2) and (1.4), the  $k$ -th order Lasserre's relaxation for (4.1) is

$$(4.4) \quad \max \quad \gamma \quad \text{s.t.} \quad f(x) - \gamma \in \Sigma_{n,2k},$$

and its dual optimization problem is

$$(4.5) \quad \min_{y \in \mathcal{M}_{2k}} \langle f, y \rangle \quad \text{s.t.} \quad M_k(y) \succeq 0, \langle 1, y \rangle = 1.$$

Clearly, for every  $k \geq d$ , (4.4) is equivalent to (4.2). However, (4.5) and (4.3) are different in satisfying flat truncation, though they have the same optimal value. Theorem 2.6 implies the following.

**Corollary 4.1.** *Let  $f_{\text{sos}}$  be the optimal value of (4.2). Suppose  $f_{\min} = f_{\text{sos}}$  and (4.1) has a nonempty set of finitely many global minimizers. Then, for  $k$  big enough, every minimizer  $y^*$  of (4.5) has a flat truncation.*

If  $y^*$  is an optimizer of (4.3) instead of (4.5), then  $y^*$  might not have a flat truncation. In this sense, (4.5) is stronger than (4.3), though their optimal values are same. If  $f_{\min} > -\infty$ , then  $f(x)$  generally has finitely many minimizers, as shown in [13, Prop. 2.1].

**4.2. Jacobian SDP relaxation.** Now we consider the Jacobian SDP relaxation for solving polynomial optimization introduced by the author in the earlier work [14]. It is a refining of relaxation (2.6). Its basic idea is to introduce some redundant polynomial equalities, say,

$$\varphi_1(x) = \cdots = \varphi_L(x) = 0,$$

which are constructed from the minors of the Jacobians of  $f(x), g_1(x), \dots, g_m(x)$ . For instance, when (1.1) has a single inequality constraint, say  $g(x) \geq 0$ , the newly introduced equalities are

$$g(x) \cdot f_{x_i}(x) = 0, \quad i = 1, \dots, n,$$

$$\sum_{1 \leq i < j \leq n, i+j=\ell} \left( f_{x_i}(x)g_{x_j}(x) - f_{x_j}(x)g_{x_i}(x) \right) = 0, \ell = 3, \dots, 2n-1.$$

When (1.1) achieves its minimum and some general nonsingularity condition (see Assumption 2.2 of [14]) holds, (1.1) is equivalent to

$$(4.6) \quad \begin{cases} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & \varphi_1(x) = \dots = \varphi_L(x) = 0, \\ & g_1(x) \geq 0, \dots, g_m(x) \geq 0. \end{cases}$$

Let  $PJ_k$  denote the  $k$ -th truncated preordering for (4.6) (cf. [14, Sec. 2.2]). Then, the resulting version of relaxation (2.6) for (4.6) is

$$(4.7) \quad \max \quad \gamma \quad \text{s.t.} \quad f(x) - \gamma \in PJ_k.$$

Similarly, its dual optimization problem is (cf. [14, Sec. 2.2])

$$(4.8) \quad \begin{cases} \min_{y \in \mathcal{M}_{2k}} & \langle f, y \rangle \\ \text{s.t.} & L_{\varphi_j}^{(k)}(y) = 0, j = 1, \dots, L, \\ & L_{g_\nu}^{(k)}(y) \succeq 0 (\forall \nu \in \{0, 1\}^m), \langle 1, y \rangle = 1. \end{cases}$$

An attractive property of the hierarchy of (4.7) is that it always has finite convergence. Thus, a practical concern in applications is: how does one identify its finite convergence? Interestingly, flat truncation could serve as a certificate for this purpose when (1.1) has finitely many minimizers.

**Corollary 4.2.** *Suppose (1.1) has a nonempty set of finitely many global minimizers and its feasible set is nonsingular (Assumption 2.2 of [14] is satisfied). Then, for all  $k$  big enough, the optimal value of (4.7) equals the global minimum of (1.1) and every minimizer of (4.8) has a flat truncation.*

*Proof.* As before, denote by  $\tilde{f}_k$  the optimal value of (4.7). By Theorem 2.3 of [14], there exists  $k_1$  such that  $\tilde{f}_{k_1}$  is equal to the global minimum of (4.6), which is also equal to  $f_{\min}$ , when Assumption 2.2 of [14] is satisfied. So, there is no duality gap between (4.7) and (4.8) when  $k$  is big. Clearly, (4.6) also has finitely many global minimizers, since it is equivalent to (1.1). Thus, the conclusion of this corollary just follows Theorem 2.6.  $\square$

In particular, if a polynomial has finitely many global minimizers in the whole space  $\mathbb{R}^n$ , then its minimizers would be obtained by using gradient SOS relaxation [11], which is a special case of Jacobian SDP relaxation (see Corollary 2.6 of [14]). Its dual version of (4.8) becomes

$$(4.9) \quad \begin{cases} \min_{y \in \mathcal{M}_{2k}} & \langle f, y \rangle \\ \text{s.t.} & L_{f_{x_j}}^{(k)}(y) = 0, j = 1, \dots, n, \\ & M_k(y) \succeq 0, \langle 1, y \rangle = 1. \end{cases}$$

Hence, we would also get the following.

**Corollary 4.3.** *Suppose a polynomial  $f(x)$  has finitely many global minimizers in  $\mathbb{R}^n$ . Then, for all  $k$  big enough, the optimal value of (4.9) equals the global minimum of  $f(x)$  in  $\mathbb{R}^n$ , and its every minimizer has a flat truncation.*

**Acknowledgement** The author was partially supported by NSF grants DMS-0757212 and DMS-0844775.

## REFERENCES

- [1] J. Bochnak, M. Coste and M-F. Roy. *Real Algebraic Geometry*, Springer, 1998.
- [2] Conway, John B. A course in Functional Analysis, Springer-Verlag, 1990, Second Edition.
- [3] R. Curto and L. Fialkow. Truncated K-moment problems in several variables. *Journal of Operator Theory*, 54(2005), pp. 189-226.
- [4] J. W. Helton and J. Nie. A Semidefinite Approach for Truncated K-Moment Problem. *Preprint*, 2011.
- [5] D. Henrion and J. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. *Positive polynomials in control*, 293310, Lecture Notes in Control and Inform. Sci., 312, Springer, Berlin, 2005.
- [6] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3): 796-817, 2001.
- [7] J.B. Lasserre, M. Laurent, and P. Rostalski. Semidefinite characterization and computation of zero-dimensional real radical ideals. *Found. Comput. Math.* 8, pp. 607–647, 2008.
- [8] J.B. Lasserre. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, 2009.
- [9] M. Laurent. Semidefinite representations for finite varieties. *Mathematical Programming*, Vol. 109, pp. 1–26, 2007.
- [10] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. *Emerging Applications of Algebraic Geometry*, Vol. 149 of *IMA Volumes in Mathematics and its Applications* (Eds. M. Putinar and S. Sullivant), Springer, pages 157-270, 2009.
- [11] J. Nie, J. Demmel and B. Sturmfels. Minimizing polynomials via sum of squares over the gradient ideal. *Math. Prog.*, Series A, Vol. 106, No. 3, pp. 587–606, 2006.
- [12] J.Nie and M. Schweighofer. On the complexity of putinar's positivstellensatz. *Journal of Complexity* 23(2007), pp.135-150.
- [13] J. Nie and K. Ranestad. Algebraic Degree of Polynomial Optimization. *SIAM Journal on Optimization*, Vol. 20, No. 1, pp. 485–502, 2009.
- [14] J. Nie. An exact jacobian SDP relaxation for polynomial optimization. *Preprint*, 2010.
- [15] P. A. Parrilo and B. Sturmfels. Minimizing polynomial functions. In S. Basu and L. Gonzalez-Vega, editors, *Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science*, volume 60 of *DIMACS Series in Discrete Mathematics and Computer Science*, pages 83-99. AMS, 2003.
- [16] M. Putinar. Positive polynomials on compact semi-algebraic sets, *Ind. Univ. Math. J.* 42 (1993), 969–984.
- [17] C. Scheiderer. Sums of squares of regular functions on real algebraic varieties. *Trans. Am. Math. Soc.*, 352, 1039-1069 (1999).
- [18] K. Schmüdgen. The K-moment problem for compact semialgebraic sets. *Math. Ann.* 289 (1991), 203–206.
- [19] M. Schweighofer. On the complexity of Schmüdgen's Positivstellensatz. *Journal of Complexity* 20, 529-543 (2004).
- [20] M. Schweighofer. Optimization of polynomials on compact semialgebraic sets. *SIAM Journal on Optimization*, Vol. 15, No. 3, pp. 805–825, 2005.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, SAN DIEGO

E-mail address: njw@math.ucsd.edu